

Mašinsko učenje - Teorija učenja

Tatjana Jakšić Krüger

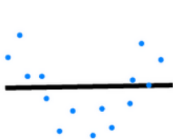
tatjana@turing.mi.sanu.ac.rs

Cilj za danas



- Formalizacija učenja.
- Pomeraj (eng. bias) i varijansa (eng. variance).
- Rešenja.

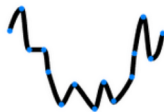
Složenost modela



Underfitting



Desired

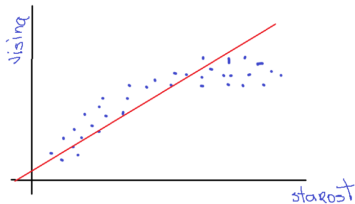


Overfitting

Izvor https://www.barelysignificant.com/slides/vendredi_quantis_2021/vendredi_quantis#32

- *overfitting* - prilagodavanje, preobučavanje.
- *underfitting* - podobučavanje.

Polinomijalna regresija

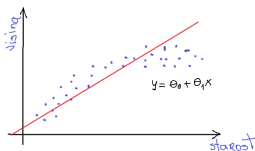


Podaci o visini i starosti za n instanci (primera).

□ Funkcija hipoteze:

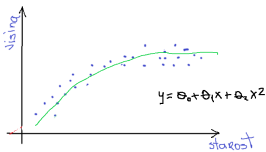
$$h(x) = \theta_0 + \theta_1x + \theta_2x^2 + \theta_3x^3.$$

Podobučavanja i preobučavanje

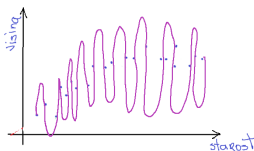


Podobučavanje.

- Visoko pristrasan, eng. *high bias*.
- Greška modela je visoka.



- Model je odgovarajući.
- Greška modela je poboljšana.



Preobučavanje.

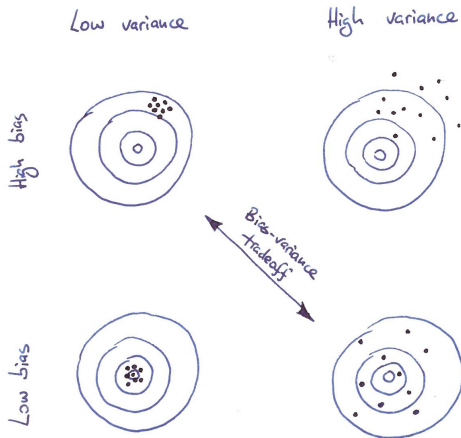
- Velika varijabilnost.
- Greška modela je nula.

Bias–variance tradeoff

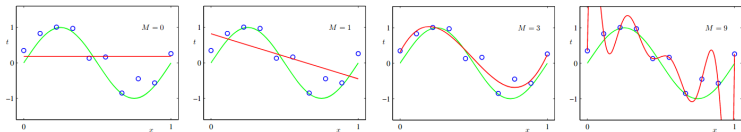
$$\begin{aligned}\text{MSE} &= \mathbb{E}\left[(y - \hat{f}(x))^2\right] = \\ &= \mathbb{E}\left[(f(x) + \epsilon - \hat{f}(x))^2\right] = \\ &= \mathbb{E}\left[(f(x) - \hat{f}(x))^2\right] + \sigma^2 = \\ &= \mathbb{E}\left[(f(x) - \mathbb{E}[\hat{f}(x)] + \mathbb{E}[\hat{f}(x)] - \hat{f}(x))^2\right] + \sigma^2 = \\ &= \underbrace{\left(f(x) - \mathbb{E}[\hat{f}(x)]\right)^2}_{\text{Bias}^2} + \mathbb{E}\left[(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2\right] + \\ &\quad + 2\left(f(x) - \mathbb{E}[\hat{f}(x)]\right)\mathbb{E}\left[\hat{f}(x) - \mathbb{E}[\hat{f}(x)]\right] + \sigma^2 = \\ &= \underbrace{\left(f(x) - \mathbb{E}[\hat{f}(x)]\right)^2}_{\text{Bias}^2} + \underbrace{\mathbb{E}\left[(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2\right]}_{\text{Variance}} + \sigma^2 = \\ &= \text{Bias}^2 + \text{Variance} + \sigma^2.\end{aligned}$$



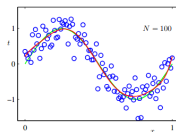
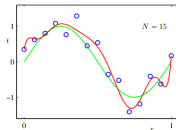
Intuition for bias and variance



Overfitting and high variance demonstration



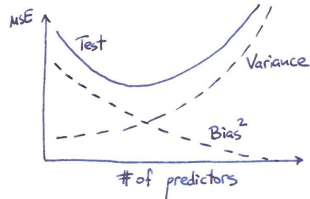
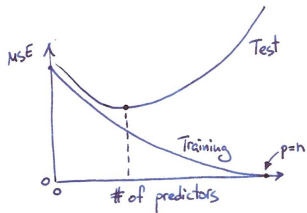
	$M = 0$	$M = 1$	$M = 3$	$M = 9$
w_0^*	0.19	0.82	0.31	0.35
w_1^*		-1.27	7.99	232.37
w_2^*			-25.43	-5321.83
w_3^*			17.37	48568.31
w_4^*				-231639.30
w_5^*				640042.26
w_6^*				-1061800.52
w_7^*				1042400.18
w_8^*				-557682.99
w_9^*				125201.43



Bishop, *Pattern Recognition and Machine Learning*



Training and test error



- Ridž regresija se zasniva na l_2 normi (euklidsko rastojanje).

$$J_{ridge}(\theta) = \frac{1}{2} \sum_{i=1}^m ((x^{(i)})^T - y^{(i)})^2 + \lambda \|\theta\|^2, \quad \lambda > 0.$$

- Laso regresija, koja se zasniva na l_1 -meri.

$$J_{lasso}(\theta) = \frac{1}{2} \sum_{i=1}^m ((x^{(i)})^T - y^{(i)})^2 + \lambda \sum_{i=1}^p |\theta_i|.$$



Hvala na pažnji.
Molim vas pitajte sve šta
vas interesuje.